MTH 4230 Lab 4 **Answer Sheet**

Due Mon., Feb. 26

1 Part A: Multiple Regression

1.1 Patient Satisfaction Data Set

1. NA
2. Print the scatterplot matrix.
3. With which **predictor** is a patient's **satisfaction** most highly correlated?

   Give the name of the predictor: _____.

   Give the value of the ***correlation***: _____.

   Which **two predictors** are most highly correlated with each other?

   Give the names of the two predictors: _____.

   Give the value of the ***correlation***: _____.

4.

   a) Write the equation of the ***fitted regression model*** $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3$ below:

   Give an **interpretation** of the value of $\boldsymbol{b_1}$.

   Give an **interpretation** of the value of $\boldsymbol{b_2}$.

   Give an **interpretation** of the value of $\boldsymbol{b_3}$.

b) Give the value of the ***coefficient of determination*** $\boldsymbol{R^2}$ (labeled `Multiple R-squared` in the output from `summary()`.

Coefficient of determination $\boldsymbol{R^2} =$ _____

Based on the value of R^2, what percentage of the variation in **satisfaction** is explained by the model with **age**, **severity**, and **anxiety**? _____

c) Look at the results of the ***model F test***, which is a test of

$$\begin{aligned} H_0: &\quad \beta_1 = \beta_2 = \beta_3 = 0 \\ H_a: &\quad \text{Not all } \beta_k\text{'s equal 0} \end{aligned}$$

Give the following values:

The observed value of the ***F* statistic** is: $F =$ _____

The p-value for the test is: **p-value** = _____

State the conclusion (Reject H_0/Fail to reject H_0). _____

Based on the result of the F test, is there a statistically significant relationship between Y and (at least one of) these predictors?

d) Give the numerical value of the estimate of the variance σ^2 of the error term ϵ in the multiple regression model. _____

Give the estimate of the standard deviation σ. _____

e) Give the following values:

The estimated standard error of b_1 is: $\boldsymbol{s\{b_1\}} =$ _____

The estimated standard error of b_2 is: $\boldsymbol{s\{b_2\}} =$ _____

The estimated standard error of b_3 is: $\boldsymbol{s\{b_3\}} =$ _____

f) Now look at the results of the ***t tests*** for individual β_k's.

Consider first the test of

$$\begin{aligned} H_0: \beta_1 &= 0 \\ H_a: \beta_1 &\neq 0 \end{aligned}$$

The observed value of the test statistic for the t test is $t =$ _____

The p-value = _____

Is the observed b_1 statistically significantly different from 0 (Yes/No)? _____

Now consider the test of

$$\begin{aligned} H_0 : \beta_2 &= 0 \\ H_a : \beta_2 &\neq 0 \end{aligned}$$

The observed value of the test statistic for the t test is $t =$ _____

The p-value = _____

Is the observed b_2 statistically significantly different from 0 (Yes/No)? _____

Finally consider the test of

$$\begin{aligned} H_0 : \beta_3 &= 0 \\ H_a : \beta_3 &\neq 0 \end{aligned}$$

The observed value of the test statistic for the t test is $t =$ _____

The p-value = _____

Is the observed b_3 statistically significantly different from 0 (Yes/No)? _____

5. ***Don't*** print the plots. Based on the plots, does the assumption of normality of the error term ϵ appear to be reasonable? (Yes/No)? _____

6. ***Don't*** print the plot. Based on the plot, does the assumption that the errors ϵ_i have a constant standard deviation appear to be reasonable? (Yes/No)? _____

2 Part B: Linearly Dependent Design Matrix

2.1 Another Hypothetical Snakes Data

1. What property of the (hypothetical) `snakes` data set prevented R from being able to estimate all the parameters in the multiple regression model?